Volume 262, number 2, 189–193 FEBS 08225 March 1990

A repeated decapeptide motif in the C-terminal domain of the ribosomal RNA methyltransferase from the erythromycin producer *Saccharopolyspora erythraea*

Namrita Dhillon* and Peter F. Leadlay

Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QW, England

Received 8 January 1990

Re-analysis of the primary structure of the ribosomal RNA *N*-methyltransferase that confers self-resistance on the erythromycin-producing bacterium *Saccharopolyspora erythraea* has confirmed the presence of a C-terminal domain containing extensive repeat sequences. Nine tandem repeats can be discerned, with a decapeptide consensus sequence GGRx(H/R)GDRRT, although no single residue is wholly invariant. This highly polar, potentially flexible domain, which is predicted to adopt either a random coil or a structure with β turns, has a counterpart in the erythromycin methyltransferase of an erythromycin-producing species of *Arthrobacter*. It also significantly resembles a portion of the C-terminal region of the eukaryotic protein nucleolin, which is unusually rich in dimethylarginine and glycine, and which is also predicted to behave as a random coil in solution. This resemblance, despite the very different roles of these proteins in ribosome biogenesis, strengthens the idea that in both rRNA methyltransferases and nucleolin these C-terminal sequences might contribute to rRNA binding.

Erythromycin; RNA-binding motif; Nucleolin; rRNA methyltransferase; Resistance gene

1. INTRODUCTION

The structural genes for the biosynthesis of the macrolide antibiotic erythromycin in *Saccharopolyspora erythraea* (formerly *Streptomyces erythraeus*) are clustered around the gene (*erm E*) that determines resistance to the antibiotic [1–5]. Resistance to erythromycin and the other 'MLS' (macrolide, lincosamide and streptogramin B) antibiotics is associated with *N*-6-dimethylation of a specific adenine residue in *S. erythraea* 23 S rRNA [6–8] which reduces the affinity of the ribosomes for the antibiotic [12]. The *erm E* gene has been cloned in *Streptomyces lividans* and *Escherichia coli* [13,14] and sequenced [15]. A comparison of the deduced amino acid sequence of the *erm E* methylase with those of the inducible *N*-methyltransferases from erythromycin-resistant bacilli and staphylococci showed that the *erm E* gene product contained a C-terminal extension with several pentapeptide RRTGG repeats [15] not present in the inducible methylases. Subsequently, the constitutive *erm A* resistance gene from the erythromycin biosynthetic cluster of a species of *Arthrobacter* [16] was also found to possess a C-terminal 'tail', similar to but shorter than that of the *erm E* methylase. We report here a re-analysis of the *erm E* methylase sequence which has suggested at least one appropriate function for the C-terminal domains in the erythromycin producers.

Correspondence address: P.F. Leadlay, Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QW, England

* *Present address:* Department of Molecular Biology, Research Institute of Scripps Clinic, 10666 North Torrey Pines Road, La Jolla, CA 92037, USA

2. MATERIALS AND METHODS

Preparation and manipulation of genomic and plasmid DNA in *E. coli*, *S. erythraea* and *Streptomyces* was carried out by established procedures [17,18]. Bacterial strains and plasmids used in this study have been described previously [5]. Protocols for DNA hybridisation and sequencing and the construction and restriction map of plasmid pRH3 have also been described [5].

DNA sequence data were compiled using the Staden programs DBAUTO and DBUTIL [19]. For sequence analysis, the Staden programs ANALYSEQ, WEIGHTS and ANALYSEPL were used together with the University of Wisconsin Genetics Computer Group programs [20]. Protein sequence alignment was also done with the program MULTALIGN [21].

3. RESULTS

3.1. *Nucleotide sequence of* erm E

No attempts were made to exclude the previously-sequenced *erm E* gene during the cloning and sequencing of a 7.3 kbp *Sac*I fragment, part of the gene cluster responsible for the biosynthesis of erythromycin in *S. erythraea* [5]. Sequencing data relevant to the results in this paper have been submitted to the EMBL nucleotide sequence database under the accession number x51891. Computer-assisted analysis using the FRAME program

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/90/$3.50

(A)

```
.....D R R L F K P V P K V D S.....
....AAGGTCGACCGCCGGCTGTTCAAGCCGGTGCCCAAGGTCGACTCG....
    >>>>>>>>>                    >>>>>>>>>
```

(B)

```
ermE        residue no. 192  DRRLFKPVPKVDS
ermA        residue no. 188  DRTLFTPVPRVHS
carB        residue no. 204  SRRWFRPVPAVDS
ermSF       residue no. 218  SRREFRPVPRVDS
                             * * *** * *
```

Fig.1. (A) The nucleotide and deduced amino acid sequence of *erm E* around the deletion, with the 9 bp direct repeat underlined, and the 33 bp deletion shown in bold. (B) Alignment [18] of the amino acid sequences of constitutive and inducible MLS resistance methylases, deduced from the respective gene sequences of *erm E* (this work); *erm A* from the erythromycin producer *Arthrobacter* sp. [13]; *car B* from the carbomycin producer *Streptomyces thermotolerans* [32]; and *erm SF* from the tylosin producer *Streptomyces fradiae* [33]. Identities are marked with an asterisk.

[22] indicated the presence of a complete ORF corresponding to the resistance methylase, although as noted previously [15] the program does not identify the last 270 nt of the ORF as coding sequence. However, despite its atypical codon usage, this region clearly encodes part of the methyltransferase, in agreement with the results of Uchiyama and Weisblum [15]. The only difference found was a small 33 bp region between codons 193 and 214 (fig.1) which is missing in the sequence reported by Uchiyama and Weisblum [15]. Perhaps significantly, the 33 bp sequence is flanked by a 9 bp direct repeat. The possibility of an insertional event during sub-cloning in M13 and pUC18 vectors during the present work was ruled out by direct sequencing of the parent plasmid pRH3 using specific oligonucleotide primers. Also, the G + C mol% content

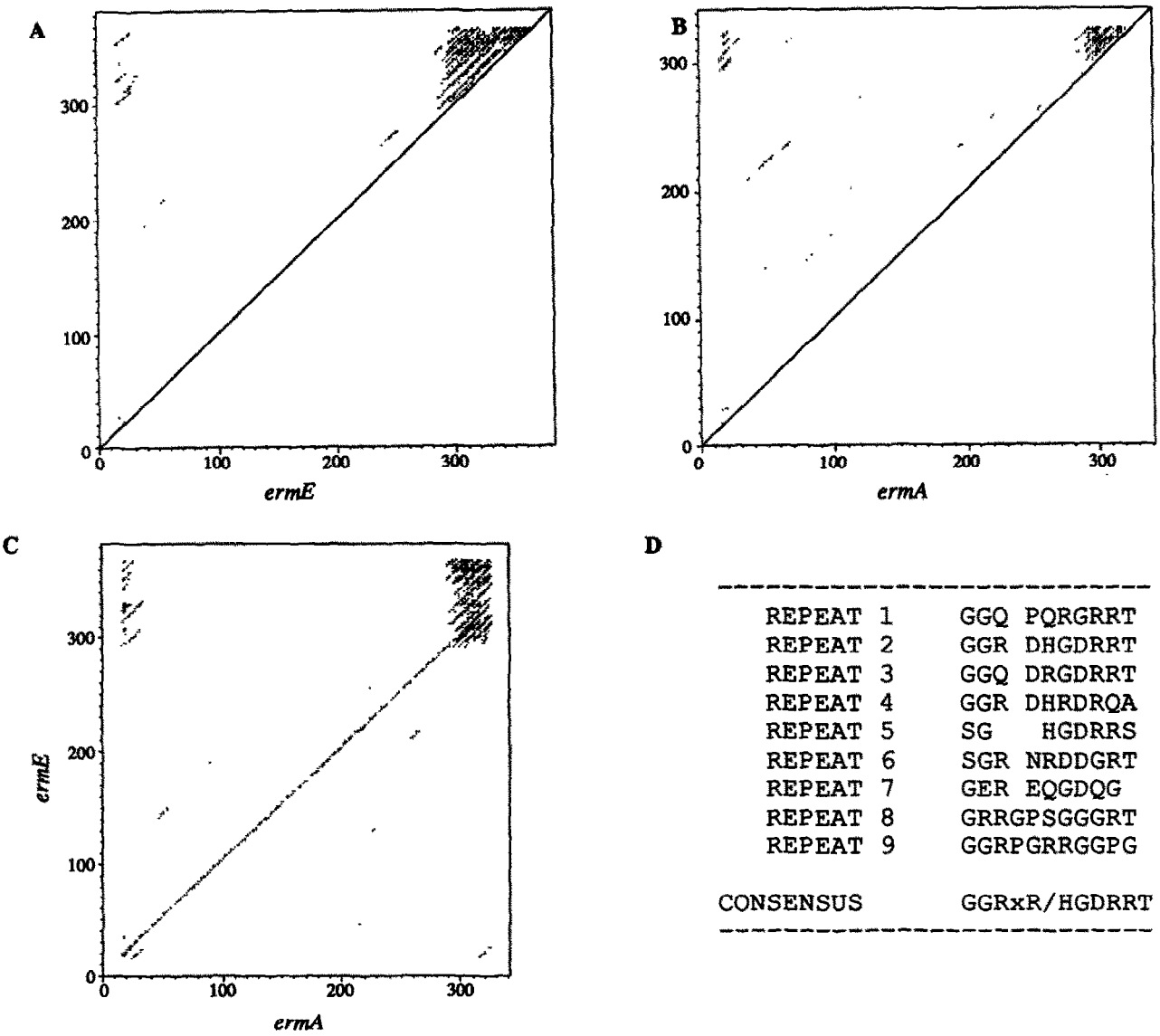

Fig.2. (A) COMPARE [16] plot of *erm E* against itself. Residues 1–381 were compared using a window of 30 and a stringency of 15. (B) COMPARE plot of *erm A* against itself. Residues 1–341 were compared as for (A). (C) COMPARE plot of *erm A* and *erm E*. Residues 1–341 and 1–381, respectively, were compared as for (A). (D) Alignment of the C-terminal repeats of *erm E* using MULTALIGN [18]. A gap penalty of 8 was used and a constant of 8 was added to the matrix file.

and codon usage of this region conforms to normal patterns found in *S. erythraea*. Finally, the insertion of this extra piece of DNA allows an even more favourable alignment to be made in this region between the amino acid sequence of the *erm E* methylase and the sequences of other erythromycin resistance methylases (fig.1).

3.2. *Analysis of C-terminal repeats in the* erm E *methylase*

Nucleotide sequencing of *erm E*, which had been previously sequenced by Uchiyama and Weisblum [15], confirmed the presence of several pentapeptide repeats at the C-terminus of the methylase. However, when the COMPARE program [22] was used to examine *erm E* self-homology at the protein level, the diagonal plot (fig.2A) showed a far more extensive stretch of repeats than expected. The C-terminal domain appears to consist of nine copies of an imperfect decapeptide tandem repeat with the consensus sequence GGRx(H/R)GDRRT [21] (fig.2D), the consensus becoming noticeably more ragged towards the C-terminus. A similar self-homology plot was obtained for the carboxy-terminal region of the *Arthrobacter* resistance gene *erm A* (fig.2B), but inspection of this sequence [16] showed no very striking consensus repeat unit. A matrix file of the *erm E* methyltransferase repeats was generated using the program WEIGHTS [19] and this file was used to screen the SWISSPROT and PIR protein sequence databanks for other potentially homologous sequences.

Apart from the *erm A* methyltransferase, this analysis revealed a significant matching score only with the C-terminal region of nucleolin, the major protein of the eukaryotic nucleolus [24,26]. The C-terminal domain of nucleolin from Chinese hamster ovary cells

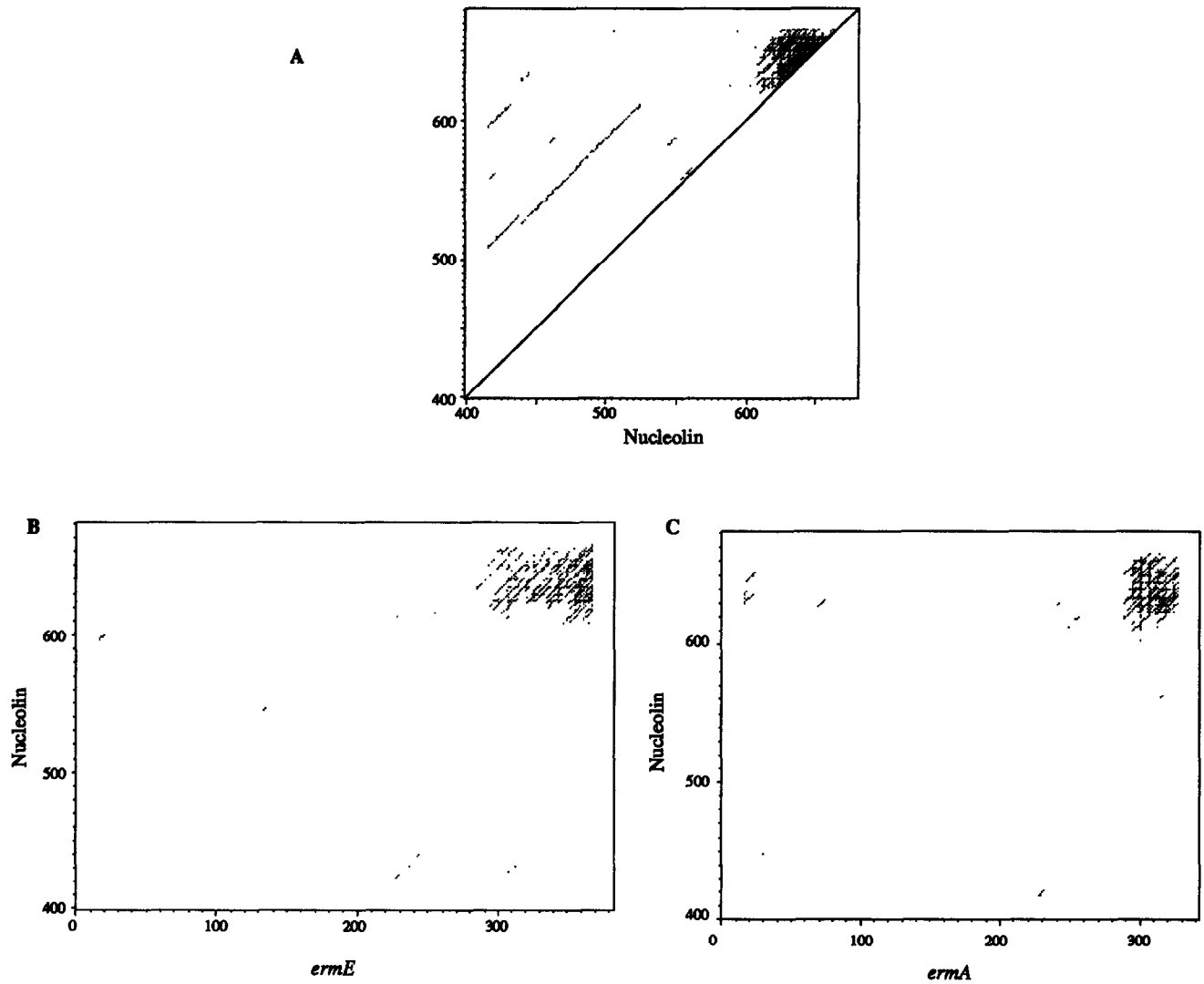

Fig.3. (A) COMPARE plot of nucleolin (residues 400–680 only) to itself. (B) COMPARE plot of *erm E* and nucleolin (residues 1–381 and 400–681, respectively). (C) COMPARE plot of *erm A* and nucleolin (residues 1–341 and 400–680, respectively). In each case the window and stringency used were as for fig.2.

[24,26] consists of a region, 52 residues long, composed largely of glycine, interspersed with 5 phenylalanine and 10 arginine residues. Most of the arginines are present as the *N,N*-dimethyl-derivative. Within this region there are significant repeats in the sequence (fig.3A), and although a consensus sequence cannot be so easily discerned as with the *erm E* methyltransferase, the decapeptide sequence GGFGGRGGGR is repeated in tandem, followed by two closely-related sequences [26]. The unexpected similarity between the *erm E* and *erm A* methylases and eukaryotic nucleolin is illustrated by the diagonal plots in fig.3B and C. Secondary structure prediction (data not shown) by the methods of Garnier [27] and Chou and Fasman [28] reinforces this similarity, in that the C-terminal domains of all 3 proteins are strongly predicted to form β-turns or random coils in solution, with little or no α-helix or extended β-structure present.

4. DISCUSSION

Re-analysis of the nucleotide sequence of the *erm E* rRNA methyltransferase gene of *S. erythraea* has confirmed that the C-terminal 90 residues constitute a highly polar, potentially flexible domain, with a high net positive charge. The gene is slightly longer than had been thought [15], because the previously-sequenced clone apparently contained a small 33 bp deletion which would easily have escaped detection. The calculated M_r of the encoded polypeptide (381 amino acids long) is 41900, and in apparent agreement with this, we have recently expressed the *erm E* gene at high levels in *Escherichia coli* and obtained a polypeptide of subunit M_r 40000, as judged by its mobility on SDS-polyacrylamide gels (Dhillon, N., Cortes, J. and Leadlay, P.F., unpublished data).

Comparison between the C-terminal amino acid sequences of the *erm E* and the *erm A* methylases shows that they do not possess any completely-conserved sequence motif, but the similarity in amino acid composition, net charge, polarity and predicted conformation in solution points to a common function for this domain. One explanation for the presence in the erythromycin producers of a basic, potentially flexible C-terminal domain is therefore that it facilitates binding of the methyltransferase to its rRNA substrate. The *erm E* gene product is active in vitro against both *S. lividans* and *E. coli* 23 S rRNA but does not act efficiently on 50 S ribosomal particles [8], and the *erm E* methyltransferase apparently co-sediments with the ribosomal fraction of *S. erythraea* [8]. However, possession of such a C-terminal tail is not essential for methyltransferase activity against free rRNA: the mostly inducible rRNA methyltransferases from Gram-positive bacteria that are not themselves producers do not contain such a C-terminal domain [9], although several of these have been shown to act on free 23 S rRNA [10,11]. The basic C-terminal 'tail' might conceivably mediate a direct interaction between the enzyme and the rRNA, or it might enhance binding through its interactions with specific ribosomal proteins.

The similarity observed between the C-terminal domains of the *erm E* and *erm A* methylases and the C-terminal domain of nucleolin owes rather more to very biased and repetitive amino acid sequence than to a common highly-conserved motif. The C-terminal 'tail' in nucleolin is also less polar, because of the regular punctuation with phenylalanine residues. However, the similarity in the overall basicity and in the predicted conformational properties of these domains remains striking. Nucleolin is a major constituent of the nucleolus in exponentially-growing eukaryotic cells, and has been implicated in the synthesis of pre-ribosomal RNA, the assembly of pre-ribosomes, and the maturation of transcripts [23,24]. Its amino acid sequence [25,26] reveals a modular arrangement of domains reflecting its multiple functions in ribosome biogenesis. Nucleolin contains 4 copies [29] of the 11–13 residue ribonucleoprotein (RNP) binding consensus [30–32], an RNA-binding motif in which a number of positively-charged residues are followed by several hydrophobic residues. Intriguingly, a similar domain arrangement has been found for heterogeneous nucleoprotein A1 (hnRNP A1), a major non-histone protein involved in the synthesis, packaging and transport of RNA. HnRNP A1 contains two copies of the RNP consensus [33] and a flexible glycine- and arginine-rich C-terminal domain. One role of such glycine- and arginine-rich domains may be to promote RNA binding by nucleating protein-protein interactions in RNP assembly [32,34], and parallels have been drawn with the non-helical glycine-rich C-terminal domains in keratins and lamins, which are important in filament assembly [32].

The primary determinants of RNA binding in nucleolin are apparently the regions containing the RNP consensus [29], but good evidence has been obtained that the C-terminal tail in hnRNP A1 may also interact directly with single-stranded nucleic acids [34]. The observation of the similarity between this domain in nucleolin and in the bacterial ribosomal *N*-methyltransferases reinforces the view that in the bacterial methylases the flexible domain contributes to rRNA binding. The ability to express the *erm E* methyltransferase efficiently from the cloned gene in *E. coli* (Dhillon, N., Cortes, J. and Leadlay, P.F., unpublished data) opens the way for a closer study of the possible role of the C-terminal 'tail' in RNA-protein recognition.

Acknowledgements: We thank Professor Eric Cundliffe, Drs Richard Hale, Jesus Cortes and Rohintan Kamakaka for useful discussions. Synthetic oligonucleotides were provided by the oligonucleotide syn-

thesis facility of the Departments of Biochemistry and Pathology, funded by the Wellcome Trust. This work was supported by an MRC project grant and by the Antibiotics and Recombinant DNA Initiative of the SERC Biotechnology Directorate. N.D. was a Cambridge Nehru Scholar. P.F.L. is a member of the SERC Cambridge Centre for Molecular Recognition.

REFERENCES

[1] Stanzak, R., Matsushima, P., Baltz, R.H. and Rao, R.N. (1986) Biotechnol. Lett. 4, 229–232.
[2] Vara, J.A., Lewandowska-Skarbek, N., Wang, Y., Donadio, S. and Hutchinson, C.R. (1990) J. Bacteriol., in press.
[3] Donadio, S., Tuan, J.S., Staver, M.J., Weber, J.M., Paulus, T.J., Maine, G.T., Leung, J.O., DeWitt, J.P., Vara, J.A., Wang, Y.-G., Hutchinson, C.R. and Katz, L. (1990) in: Genetics and Molecular Biology of Industrial Microorganisms (Hershberger, C.L., Queener, S.W. and Hegeman, G. eds) American Society of Microbiology, Washington, DC, in press.
[4] Weber, J.M., Schoner, B. and Losick, R. (1989) Gene 75, 235–241.
[5] Dhillon, N., Hale, R.S., Cortes, J. and Leadlay, P.F. (1989) Mol. Microbiol. 3, 1405–1414.
[6] Teraoka, H. and Tanaka, K. (1974) J. Bacteriol. 120, 316–321.
[7] Graham, M.-Y. and Weisblum, B. (1979) J. Bacteriol. 137, 1464–1467.
[8] Skinner, R.H., Cundliffe, E. and Schmidt, F.J. (1983) J. Biol. Chem. 258, 12702–12706.
[9] Kamimiya, S. and Weisblum, B. (1988) J. Bacteriol. 170, 1800–1811.
[10] Zalacain, M. and Cundliffe, E. (1989) J. Bacteriol. 171, 4254–4260.
[11] Jenkins, G., Zalacain, M. and Cundliffe, E. (1990) J. Gen. Microbiol., in press.
[12] Dubnau, D. (1984) CRC Crit. Rev. Biochem. 16, 103–132.
[13] Bibb, M.J., Jannsen, G.R. and Ward, J.M. (1986) Gene 41, E357–E368.
[14] Katz, L., Brown, D., Boris, K. and Tuan, J. (1987) Gene 55, 319–325.
[15] Uchiyama, H. and Weisblum, B. (1985) Gene 38, 103–110.
[16] Roberts, A.N., Hudson, G.S. and Brenner, S. (1985) Gene 35, 259–270.
[17] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[18] Hopwood, D.A., Bibb, M.J., Chater, K.F., Bruton, C.J., Kieser, H.M., Lydiate, D.J., Smith, C.P., Ward, J.M. and Schrempf, H. (1985) Genetic Manipulation of Streptomyces – A Laboratory Manual, John Innes Foundation, Norwich, England.
[19] Staden, R. (1984) Nucleic Acids Res. 12, 521–528.
[20] Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 387–395.
[21] Barton, G.J. and Sternberg, M.J.E. (1987) J. Mol. Biol. 198, 353–354.
[22] Bibb, M.J., Findlay, P.R. and Johnson, M.W. (1984) Gene 30, 157–166.
[23] Herrera, A.H. and Olson, M.O.J. (1986) Biochemistry 25, 6258–6263.
[24] Bourbon, H.M., Bugler, B., Caizergues-Ferrer, M., Amalric, F. and Zalta, J.P. (1983) Mol. Biol. Rep. 9, 39–47.
[25] Lapeyre, B., Amalric, F., Ghaffari, S.H., Venkatarama Rao, S.V., Dumbar, T.S. and Olson, M.J. (1986) J. Biol. Chem. 261, 9167–9173.
[26] Lapeyre, B., Bourbon, H. and Amalric, F. (1987) Proc. Natl. Acad. Sci. USA 84, 1472–1476.
[27] Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) J. Mol. Biol. 120, 97–120.
[28] Chou, P.Y. and Fasman, G.D. (1974) Biochemistry 13, 222–245.
[29] Bugler, B., Bourbon, H., Lapeyre, B., Wallace, M.O., Chang, J.-H., Amalric, F. and Olson, M.J. (1987) J. Biol. Chem. 262, 10922–10925.
[30] Adam, S.A., Nakagawa, T., Swanson, M.S., Woodruff, T.K. and Dreyfuss, G. (1986) Mol. Cell. Biol. 6, 2932–2943.
[31] Merrill, B.M., LoPresti, M.B., Stone, T.L. and Williams, K.R. (1986) J. Biol. Chem. 216, 878–883.
[32] Chung, S.Y. and Wooley, J. (1986) Proteins 1, 195–210.
[33] Cobianchi, F., SenGupta, D.N., Zmudzka, B.Z. and Wilson, S.H. (1986) J. Biol. Chem. 261, 3536–3543.
[34] Cobianchi, F., Karpel, R.L., Williams, K.R., Notario, V. and Wilson, S.H. (1987) J. Biol. Chem. 263, 1063–1071.